1 Probability

1.1 Random Variables and Independence

A *real-valued random variable* is a variable that can take each of a set of possible values in $\mathbb{R}$ with a certain probability.

Two random variables, X and Y are said to be *independent* if $P(X = x|Y = y) = P(X = x)$ for all values x of X and y of Y.

1.2 Probability Distribution Function (PDF)

The *probability distribution of a real-valued random variable X* specifies the probability that X takes each of its possible values.

For example, suppose you flip a coin two times. This simple statistical experiment can have four possible outcomes: HH, HT, TH, and TT. Now, let the variable X represent the number of Heads that result from this experiment. X is a *random variable* that can take the values 0, 1, and 2. The table below, which associates each outcome with its probability, is the probability distribution of the discrete random variable X:

Number of heads	Probability
0	0.25
1	0.50
2	0.25

1.3 Cumulative Probability Distributions (CDF)

The *cumulative probability distribution of a real-valued random variable X* specifies, for each of the possible values x that X can take, the probability that X takes a value equal to or smaller than x.

Let us return to the coin flip experiment. If we flip a coin two times, we might ask: What is the probability that the coin flips would result in one or fewer heads? It would be the probability that the coin flip experiment results in zero heads plus the probability that the experiment results in one head.

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.25 + 0.50 = 0.75$$

The table below gives the cumulative probability distribution of the discrete random variable X:

Number of heads	Probability $P(X = x)$	Cumulative Probability $P(X < x)$
0	0.25	0.25
1	0.50	0.75
2	0.25	1.00

Sanity Test: Are PDF's guaranteed to be non-decreasing? What about CDF's?

Solution: PDF's are not guaranteed to be non-decreasing (for example, the PDF for any normal distribution is not non-decreasing). In contrast, CDF's are always non-decreasing.[1]

1.4   The Mean of a Discrete Random Variable

The mean or expectation of a real-valued random variable X is a weighted average of the possible values that the random variable can take. For example, if the random variable X can take values in a finite[2] set $A \subset \mathbb{R}$, its mean $E(X)$ is defined by

$$E(X) := \sum_{x \in A} P(x)x$$

Unlike the sample mean of a group of observations, which gives each observation equal weight, the mean of a random variable weights each outcome x according to its probability, $P(x)$. The mean of a random variable provides the long-run average of the variable, or the expected average outcome over many observations.

For example, suppose an individual plays a gambling game where it is possible to lose \$1, break even, win \$3 or win \$10 each time she plays. Let Y be the random variable that takes the value $y \in \{-1, 0, 3, 5\}$ when the outcome of the gamble is y. The probability distribution of Y is provided by the following table:

y	Probability
-1	.3
0	.4
3	.2
5	.1

Calculate the expectation of Y.

Solution: The mean of the random variable Y can be calculated as follows:
$$\begin{aligned} \mu(Y) &= (-1 \times .3) + (0 \times .4) + (3 \times .2) + (5 \times .1) \\ &= -0.3 + 0.6 + 0.5 \\ &= 0.8 \end{aligned}$$

---

[1] To see this, consider that the CDF is a sum (or integral) over an interval of the PDF; specifically, for a discrete random variable defined over (a, b), $P(X \leq x) = \sum_{i=a}^{x} P(X = x)$ and for a continuous random variable, $P(X \leq x) = \int_{-\infty}^{x} f(x)$ where $f(x)$ is the PDF of X. PDFs are nonnegative, so $x_1 < x_2$ implies $P(X \leq x_1) \leq P(X \leq x_2)$, since $f(x_2) = P(X = x_2) \geq 0$; therefore, CDFs are non-decreasing.

[2] If A is infinite, we replace sums for integrals.

In the long run, then, the player can expect to win about 80 cents playing this game – the odds are in her favor.

Properties of Expectation

- For *any* two random variables X and Y, $E(X + Y) = E(X) + E(Y)$.

- For two *independent* random variables X and Y, $E(XY) = E(X)E(Y)$.

1.5 The Variance of a Discrete Random Variable

The *variance* of a real-valued random variable X measures the spread, or variability, of the distribution of X. For example, suppose X can take values in a finite[3] set $A \subset \mathbb{R}$. Then the variance $\mathrm{Var}(X)$ of X is defined by

$$\mathrm{Var}(X) := \sum_{x \in A} P(x)(x - E(X))^2$$

For example, in the gambling game above, the variance of the random variable Y may be calculated as follows:

$$
\begin{aligned}
\mathrm{Var}(Y) &= (-1 - 0.8)^2 \times 0.3 + (0 - 0.8)^2 \times 0.4 + (3 - 0.8)^2 \times 0.2 + (5 - 0.8)^2 \times 0.1 \\
&= (-1.8)^2 \times 0.3 + (-0.8)^2 \times 0.4 + (2.2)^2 * 0.2 + (4.2)^2 \times 0.1 \\
&= 3.24 \times 0.3 + 0.64 \times 0.4 + 4.84 \times 0.2 + 17.64 \times 0.1 \\
&= 0.972 + 0.256 + 0.968 + 1.764 = 3.960
\end{aligned}
$$

Hence, the variance and standard deviation of Y are 3.96 and $\sqrt{3.96} = 1.99$ respectively.

Unlike the sample variance of a group of observations, which gives each observation equal weight, the variance of a random variable weighs each outcome x according to its probability, $P(x)$. The *standard deviation* of the random variable X is the square root of its variance and is sometimes denoted by σ.

Note that standard deviation has a more natural interpretation than variance. This is because variance is measured in squared units, whereas standard deviation is in the same units as those in which the random variable itself is measured.

Properties of Variance   For *independent* random variables X and Y:

- $\mathrm{Var}(X) \geq 0$

- $\mathrm{Var}(X) = E(X^2) - E(X)^2$

- $\mathrm{Var}(X + Y) = \mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

---

[3]If A is infinite, we replace sums for integrals.

2 Examples of Common Probability Distributions

2.1 Uniform Probability Distribution

The simplest probability distribution occurs when all of the values of a random variable occur with equal probability. This probability distribution is called the *uniform distribution*.

For example, suppose that a fair die is tossed. Let X be the random variable that takes the value x when the die lands on x, for $x \in \{1, 2, 3, 4, 5, 6\}$. Since the probability that a fair die lands on each the possible numbers is the same, the random variable X has a uniform distribution.

2.2 Binomial Distribution

A *Bernoulli trial* is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success p is the same every time the experiment is conducted.

Closely related to a Bernoulli trial is a *binomial experiment*, which consists of a fixed number of statistically independent Bernoulli trials, each with a probability of success p, and counts the number of successes. A random variable corresponding to a binomial experiment with n Bernoulli trials is denoted by $B(n, p)$, and is said to have a binomial distribution. The probability of exactly k successes in the binomial experiment is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\binom{n}{k} := \frac{n!}{(n-k)!k!}$, which reads "$n$ choose k", is the number of different sets of k elements we can choose out of a set of n elements.

For example, suppose we create edges between each pair of a set nodes of nodes $\{a, b, c, d\}$ with probability p. The random variable "number of edges in the network" is distributed according to a binomial distribution $B(n, p)$.

2.2.1 Poisson Distribution

The random variable X has a *Poisson distribution with parameter* $\lambda > 0$ if, for $k = 0, 1, 2, \ldots$:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

The positive real number λ is equal to the mean and variance of the random variable X. That is, $\mu(X) = \lambda$ and $\text{Var}(X) = \lambda$.

The Poisson distribution can be derived as a limiting case of the binomial distribution as the number of trials goes to infinity and the expected number of successes remains fixed. Therefore it can be used as an approximation to the binomial distribution if n is sufficiently large and p is sufficiently small. There is a rule of thumb stating that the Poisson distribution with $\lambda = np$ is a good approximation of the binomial distribution $B(n, p)$ if n is at least 20 and p is smaller than or equal to 0.05, and an excellent approximation if n is at least 100 and np is smaller than 10.
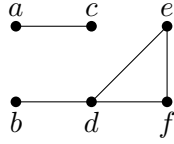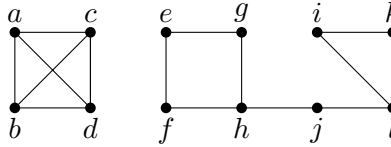
Figure 1: graph A

Figure 2: Graphgraph B

3   The Exponential and Logarithm Functions

The exponential function e^x can be defined in various ways that provide useful formulas. Fixing $x \in \mathbb{R}$,

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

The identity above will frequently prove useful in the analysis of networks. When you see an expression like $\left(1 + \frac{x}{n}\right)^n$, remember that you can approximate this expression by e^x when n is "large".

Another definition of e^x is given by

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

The logarithm function $\log(x)$ is the inverse of the exponential function e^x. That is, for all x we have $\log(e^x) = x$ and for all $x > 0$ we have $e^{\log(x)} = x$.

4   Graph Theory

4.1   Graph representations .

(Jackson 2.1.1 and 2.1.2) A graph $G = (V, E)$ is a set of *nodes* V and a set of node pairs E, called *edges*. An *undirected* graph is a graph in which the existence of an edge from a node u to a node v implies the existence of an edge from v to u. A *directed* graph is a graph in which the existence of an edge from u to v does not necessarily imply the existence of an edge from v to u. Unless stated otherwise, assume the graph is undirected. There are three ways to represent a graph:

- *Picture*: A graph can be represented via a picture. For example, Figures 1 and 2 represent

two different graphs[4].

- *Adjacency matrix*: A graph can be represented by listing its nodes $\{1, 2, \ldots, n\}$ and a real-valued $n \times n$ matrix g, where g_{ij} equals 1 if there is an edge between i and j and equals 0 otherwise[5]. For example, graph A in Figure 1 can be represented by listing its nodes $\{a, b, c, d, e, f\}$ and the adjacency matrix:

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

- *List of nodes and edges*: A graph can be represented by listing its nodes and edges. For example, graph A can be represented as a list:[6]

$$(\{a, b, c, d, e, f\}, \{ac, bd, de, df, ef\})$$

4.1.1   Exercises

a. Represent graph B in Figure 2 via its adjacency matrix and via the list of its nodes and edges.

b. Graph C has nodes $\{1, 2, 3, 4, 5, 6\}$ and the adjacency matrix:

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Represent graph C via a picture and via the list of its nodes and edges[7].

c. The representation of graph D in terms of a list of its nodes and edges is:

$$(\{a, b, c, d\}, \{ab, ac, ad, bc, bd, cd\})$$

Represent graph D via a picture and via its adjacency matrix.

Solution:

---

[4]Figures are located in the last page of this document.

[5]A *weighted* graph is one whose adjacency matrix contains numbers other than 0 and 1, but for now we will focus on unweighted graphs.

[6]Convention: for any objects $x_1, x_2, \ldots, x_n$, we denote by $\{x_1, x_2, \ldots, x_n\}$ the set that contains the elements $x_1, x_2, \ldots, x_n$ and we denote by $(x_1, x_2, \ldots, x_n)$ the *ordered* set (or list) that contains x_1 as its first element, x_2 as its second element, and x_i as its ith element. The distinction is important. For example, we have $\{x_1, x_2\} = \{x_2, x_1\}$, but in general $(x_1, x_2) \neq (x_2, x_1)$.

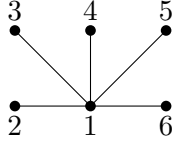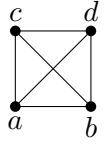[7]This type of graph is sometimes called "star graph".

Figure 3: Network C

Figure 4: Network D

a. Adjacency matrix:

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

List of nodes and edges:

$$(\{a, b, c, d, e, f, g, h, i, j, k, l\}, \{ab, ac, ad, bc, bc, cd, ef, eg, fh, gh, hj, ik, il, jl, lk\})$$

b. Picture: See Figure 3.

List of nodes and edges:

$$(\{1, 2, 3, 4, 5, 6\}, \{12, 13, 14, 15, 16\})$$

c. Picture: See Figure 4.

Adjacency matrix:

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

4.2 Subgraphs.

A *subgraph* of a given graph G is a graph that can be obtained from G by deleting some of its nodes and edges. For example, the graph $(\{a, c\}, \{ac\})$ is a subgraph of graph A, but the graph

$(\{a, b, c\}, \{bc\})$ is not. Represent a subgraph of graph B via a picture, via its adjacency matrix, and via the list of its nodes and edges.

4.3   Paths and Cycles

(Jackson 2.1.3)

- A *walk* in a graph G between nodes i and j is a sequence of edges in G, say $(i_1 i_2, i_2 i_3, \ldots, i_{k-1} i_k)$, with $i_1 = i$ and $i_k = j$.

- A *path* in a graph G between nodes i and j is a sequence of edges in G, say $(i_1 i_2, i_2 i_3, \ldots, i_{k-1} i_k)$, with $i_1 = i$ and $i_k = j$, and such that each node in the sequence $(i_1, i_2, \ldots, i_k)$ is distinct.

- A *cycle* is a walk that starts and ends at the same node, with no node appearing more than once except the starting node, which also appears as the ending node. For example, (uv, vw, wu) is a cycle in graph A.

- A *geodesic* between two nodes is a shortest path between them; that is, a path with no more edges than any other path between these nodes. For example, in graph A, (df) is a geodesic between nodes d and f, but (de, ef) is not.

4.4   Components and Connected Subgraphs.

(Jackson 2.1.5) A graph is *connected* if every two nodes are connected by some path. For example, graph A is not connected, but its subgraph

$$(\{b, d, e, f\}, \{bd, df, de, ef\})$$

is connected. A graph is *completely connected* if it is has an edge between every two nodes.

A *component* C of graph G is a (nonempty) subgraph that (i) is connected and that (ii) is maximal – i.e. it is such that all other connected subgraphs of G with some node in common with C are subgraphs of C.[8] For example, both $(\{a, c\}, \{ac\})$ and $(\{b, d, e, f\}, \{bd, de, df, ef\})$ are components of graph A, but $(\{d, e, f\}, \{de, df, ef\})$ is not, since it is a subgraph of the component $(\{b, d, e, f\}, \{bd, df, de, ef\})$ and hence not maximal.

a. Write down all the components of graph B via a list of their nodes and edges.

Solution:

a.
$$(\{a, b, c, d\}, \{ab, ac, ad, bc, bd, cd\})$$

and
$$(\{e, f, g, h, i, j, k, l\}, \{ef, eg, fh, gh, hj, ik, il, jl, kl\})$$

---

[8]Intuitively, a component is a piece of the graph not connected to anything else.

4.5 Neighborhood.

(Jackson 2.1.7) The *neighborhood of a node i* is the set of nodes that i is linked to. For example, the neighborhood of the node f in the graph A is $\{d, e\}$. Given a set of nodes S, the *neighborhood of S* is the union of the neighborhoods of all of its members – that is, the set of nodes that have at least one edge to some node in S. For example, the neighborhood of the nodes $\{a, b\}$ in graph A is $\{c, d\}$.

  a. What is the neighborhood of node d in graph B?

  b. What is the neighborhood of the set of nodes $\{e, g\}$ in graph B?

Solution:

  a. $\{a, b, c\}$

  b. $\{e, g, f, h\}$. Note that $\{e, g\}$ is a subset of the neighborhood of $\{e, g\}$ since e neighbors g and vice versa.

4.6 Degree and Graph Density.

(Jackson 2.1.8) The *degree* of a node is the number of links that involve that node. For example, the degree of the nodes a and d in graph A are 1 and 3 respectively.

In graph B:

  a. What is the degree of node a?

  b. What is the average degree of all the nodes in the graph?

  c. What is the average degree of node a's neighbors?

Solution:

  a. 3

  b. In alphabetical order, each nodes degree is 3,3,3,3,2,2,2,3,2,2,2,3. The average degree is then 2.5.

  c. The average degree of a's neighbors is 3.

5 Big-O Notation

5.1 Formal Definition

Big-O notation is mathematical notation used to describe the behavior of a function as their arguments approach infinity. Formally, $O(\cdot)$ is defined as follows: $f(x) = O(g(x))$ as $x \to \infty$ if there

exists a positive constant M such that for all $x \geq x'$, $|f(x)| \leq M|g(x)|$. Using this definition, we can see that two linear functions, e.g. $f(x) = 5x$ and $g(x) = x$, follow the relationship $f(x) = O(x)$ because for $M = 5$ the definition holds. In fact, for any monomial $f(x) = ax^b$ ($x^7, 6x, 9$, etc.), $f(x) = O(x^b)$.

It is useful to think of $O(\cdot)$ as an inequality, i.e. $g(x)$ upper bounds $f(x)$: $f(x)$ "$\leq$" $g(x)$ in the limit. Then we denote "$\geq$" with Ω: $f(x) = \Omega(g(x))$ if there exists a positive constant M such that for all $x \geq x'$, $|f(x)| \geq M|g(x)|$. For strict inequalities, We say that $f(x) = o(g(x))$ (small o) if $|f(x)| < M|g(x)|$ for $x \geq x'$ and similarly for ω. Finally, we say that $f(x) = \Theta(g(x))$ if $f(x) = O(g(x))$ and $f(x) = \Omega(g(x))$, which can be thought of as "equality".

5.2   In Practice

In practice, it is cumbersome to deal with the definition of big-O notation, so we employ a number of rules to simplify our analysis:

- if $f(x)$ is a sum of terms, we only need to consider the term with the largest growth rate. For polynomials, this means only considering the term with the highest exponent.

- if $f(x)$ is a product of terms, we can ignore multiplicative constants

5.3   The Growth of Some One-Variable Functions

Given two increasing functions, $f, g : \mathbb{R} \to \mathbb{R}$ we are interested in answering the question of which of the two functions is larger for "large x". Here we will provide some illustrations using the functions

- x (yellow),

- $x + 5$ (magenta),

- $x * 5$ (cyan),

- x^5 (red),
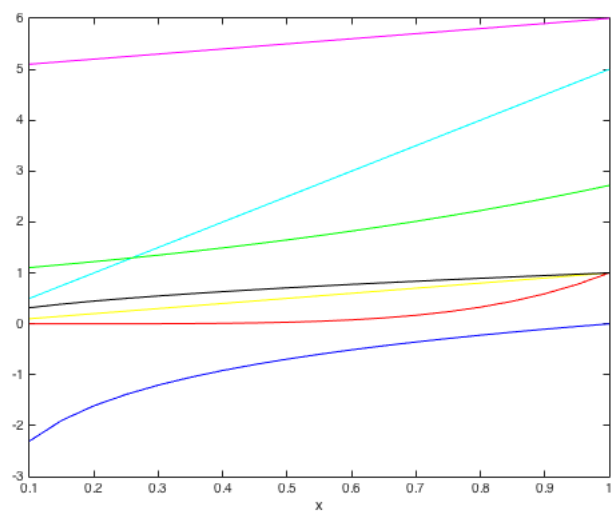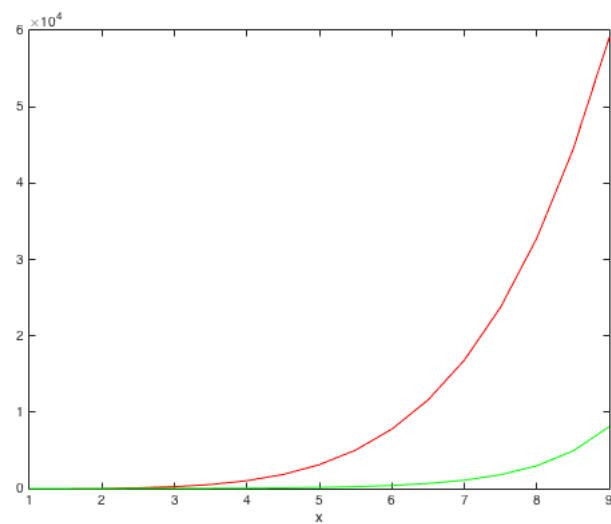
- e^x (green),

- $log(x)$ (blue) and

- $\sqrt{x}$ (black).
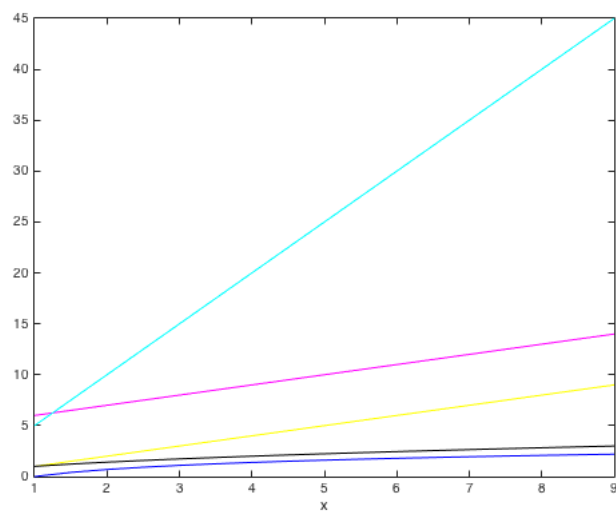
Figure 5:
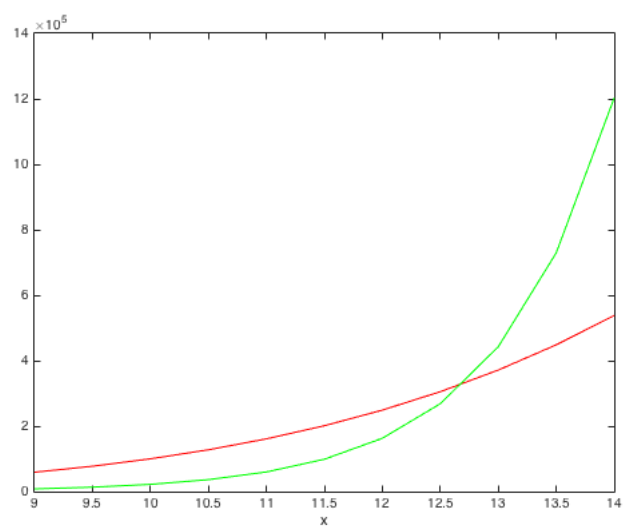

Figure 6:

Figure 7:

Figure 8:

Figure 9: